Enhanced Method for Digital Projection of Root Carving Paper Cultural Relics: Reconstruction and Rendering Optimization Based on 3D Gaussian Splatting

Youming Wang,[a,b] Nan Zhou,[a] Zhixiong Zhang,[c] Youwu Xu,[a] Zeling Li,[a] and Wanhui Gao [a,*]

This study proposes a 3D reconstruction method based on 3D Gaussian Splatting (3DGS) for cultural relics, such as root carvings and paper-based relics, which are characterized by fine fiber structures and complex surface details. A consumer-grade smartphone was used to capture multi-view images of the root carvings and paper-based relics. Subsequently, the FFmpeg tool was employed to extract image frames, which serve as input images. COLMAP software was utilized to perform feature matching, Structure from Motion (SfM) computation, and camera pose estimation. This process generated a sparse point cloud, which was used as the initialization data for the 3D Gaussian distribution. This workflow produced high-quality reconstruction of the micro-details of paper materials (such as fiber textures and carved indentations). Comparative evaluations were conducted against Neural Radiance Fields (NeRF) and Instant Neural Graphics Primitives (Instant-NGP). The proposed 3DGS framework achieved superior performance. Compared with NeRF, the average PSNR increased by 39.60%, SSIM by 65.84%, and LPIPS by 90.57%, while the reconstruction time was shortened by 68.89%. Compared with Instant-NGP, the framework achieved an increase of 31.56% in average PSNR, 42.58% in SSIM, a decrease of 89.21% in LPIPS, and a reduction of 48.15% in reconstruction time.

Contact information: a: Zhejiang College of Construction, Hangzhou, 310000, China; b:University of Shanghai for Science and Technology, Shanghai,200000,China; c: Robotic Brain Perception Department, Hangzhou Zhanzo Technology Co., Ltd. Hangzhou, 310000, China;
* Corresponding author: gwh20181028@163.com

INTRODUCTION

In the field of cultural heritage conservation, digital modeling of cultural relics enables the permanent archiving and virtual display of endangered relics. Among these, root carvings and ancient books, which are vital components of world cultural heritage, integrate exquisite carving art and traditional papermaking techniques, respectively. However, as root carvings and ancient books primarily consist of natural organic materials, they are highly susceptible to environmental factors such as humidity, temperature, and light. Long-term storage can lead to deterioration, including fiber degradation, surface cracking, and structural damage. Digital technology, capable of converting physical relics into high-precision 3D models, has become a key means of protecting cultural assets. It facilitates the long-term archiving of relics, virtual exhibitions, and scientific research in the field of cultural heritage (Gomes *et al.* 2014; Dell'Unto *et al.* 2018; Munoz-Pandiella *et al.* 2024).

Traditional 3D modeling methods based on laser scanning and structured light can obtain high-precision point cloud data; however, they suffer from drawbacks such as bulky equipment, distorted texture mapping, and limitations in capturing details of complex curved surfaces (Jebur 2022). Image-based 3D reconstruction aims to recover the 3D structure and geometric shape of a target object or scene from a set of input images (Samavati and Soryani 2023). Given that root carvings and ancient books exhibit semi-transparent properties and fine microfiber structures, Structure from Motion-Multi-View Stereo (SfM-MVS) technology has emerged as a powerful tool for capturing high-fidelity 3D models of such objects (Zhao *et al.* 2021). By utilizing multiple overlapping images to reconstruct an object's geometry and texture, SfM-MVS is particularly suitable for modeling complex surfaces and fine details (Moyano *et al.* 2023). Nevertheless, traditional SfM-MVS methods often struggle with semi-transparent materials, as their light refraction properties can cause distortions in depth estimation and texture mapping (Ryan *et al.* 2024). To address this issue, recent advancements in Neural Radiance Fields (NeRF) have demonstrated potential in handling semi-transparent materials, by more accurately simulating light interaction processes.

Variants such as Mip-NeRF (Barron *et al.* 2021) and Mip-NeRF360 (Barron *et al.* 2021) have optimized NeRF's performance through multi-scale representation and anti-aliasing optimization techniques. In parallel, methods such as FastNeRF (Garbin *et al.* 2021) and R2L (Wang *et al.* 2022) focus on enhancing NeRF's efficiency *via* diverse technical approaches, significantly improving both training speed and real-time rendering capabilities. However, NeRF still requires substantial computational resources for scene optimization, and it suffers from blurry rendering results and excessive rendering time. NeRF++ (Zhang *et al.* 2020) introduced an inverted sphere parameterization scheme, which places the main reconstruction scene inside a unit sphere. For regions outside the sphere, it normalizes the distance parameter R to $1/R$, thereby effectively alleviating the blurriness issue in reconstruction. Mega-NeRF (Turki *et al.* 2021) adopts the core strategy of NeRF++ and introduces targeted improvements to address large-scale scene training. It divides the scene into independent regions and trains each region using a separate NeRF model. To address training latency, Instant-NGP (Müller *et al.* 2022) replaces the positional encoding of the original NeRF with hash encoding based on multi-resolution grids. It models the scene as grids of varying resolutions while preserving the principles of ray sampling and point sampling, and it encodes sampling points within the grids as hash table indices corresponding to trainable feature vectors. By using a smaller neural network, Instant-NGP accelerates both training and rendering. These technologies not only improve the accuracy of 3D reconstruction but also provide more immersive and realistic digital representations of cultural relics, which is crucial for conservation and dissemination efforts (Nobuhara and Matsuyama 2003). Despite these advancements, challenges remain in the digital conservation of cultural relics. A major issue is NeRF's computational complexity. The generation of high-quality models requires significant computing power and time, which is particularly problematic for large-scale digitalization projects involving numerous relics. Additionally, NeRF methods often lack the high-frequency fidelity required to capture fine details of paper fibers, and these shortcomings become more pronounced when handling subtle textural variations inherent in natural material fiber structures.

3D Gaussian Splatting (3DGS) uses a Gaussian distribution-based approach to representing surface details, offering high fidelity in capturing micro-fiber structures and fine textures. This technology models scenes using Gaussian functions and sparse point

bioresources.cnr.ncsu.edu

cloud data, converting 3D information into the form of Gaussian functions to enable scene model reconstruction and real-time rendering (Rhodin *et al*. 2015; Kerbl *et al.* 2023; Yang *et al.* 2024). Its improved algorithm, Mip-Splatting (Yu *et al.* 2024), addresses issues of low resolution and distant camera poses by proposing to represent the same scene using Gaussian ellipsoids of different scales. This approach enhances rendering speed and model quality by controlling the size of these ellipsoids. For large-scale scene reconstruction, Do Gaussian (Chen and Lee 2024) first proposes decomposing the scene to be reconstructed into multiple overlapping regions. Then it performs distributed training on these regions using the alternating direction method of multipliers. In this way it ensures the stability of global Gaussians by sharing Gaussian features across different regions. 3D GS technology and its improved algorithms offer new possibilities for the 3D modeling of complex geometric structures such as root carvings. The advent of 3DGS has effectively addressed the problem of high computational load in NeRF's rendering and training processes, enabling real-time high-quality 3D reconstruction. Wang (2024) proposed a Gaussian Splatting method based on continuous close-range and long-range photography using Unmanned Aerial Vehicles (UAVs), which achieves high-precision real-time rendering and reconstruction of relics and buildings within 20 minutes. Zheng *et al.* (2025) reconstructed dynamic street scenes based on 3D Gaussian point distribution technology, achieving high-fidelity dynamic scene reconstruction and moving object removal. Cai *et al.* (2024) used 3DGS for urban building reconstruction, improving training speed and reconstruction quality by optimizing and accelerating rendering algorithms, thus providing a new solution for urban building 3D reconstruction. In the field of cultural relics research, 3DGS holds significant potential in 3D modeling of cultural relics due to its potential in high-quality visual fidelity and accurate geometric representation.

This study proposes a 3DGS-driven framework specifically tailored for root carvings and paper-based cultural relics. The core contributions of the study include designing a low-cost data acquisition workflow based on consumer-grade smartphones, which prevents damage to fragile paper-based relics while reducing equipment costs. The method improves 3DGS optimization strategies-including adaptive Gaussian density control and micro-detail preservation and accurately captures fiber textures and carved contours of root carvings and paper-based relics. The goal is to systematically verify the framework's performance through comparative experiments with NeRF and Instant-NGP, thereby providing technical guidelines for the digital preservation of cultural heritage.

The core characteristics of root carving and paper-based relics, such as their dense fiber interweaving and non-uniform texture distribution, present specific challenges for 3D reconstruction. Paper-based relics possess a degree of translucency, while the wooden fibers in root carvings exhibit subtle light scattering. The standard density control mechanism in the original 3DGS framework is inadequate for adapting to these material properties, often leading to reconstruction artifacts such as textural blurring and uneven light transmission. The present method addresses this by optimizing the density control, aligning the distribution of Gaussian points with the material's light-transmitting characteristics. This ensures authentic texture restoration in semi-translucent areas and enhances the overall realism and accuracy of the reconstruction.

The proposed method is not limited to the digital preservation of bio-based cultural relics such as root carvings and paper-based relics. It holds significant potential for extension to other domains involving intricate biological structures, such as wooden bio-based building materials, plant fiber-based cultural and creative products, and bio-based archaeological specimens. This approach provides a unified solution to the common

challenges of accurately reconstructing fine structural details while addressing the preservation needs of fragile bio-based materials across different application scenarios.

EXPERIMENTAL

Materials and Data

The experimental dataset comprised 300 root carving and paper-based relics sourced from the Zhejiang Zhijiang Museum. These relics were classified based on their morphological characteristics, geometric structures, and texture features, resulting in three categories for root carvings and paper-based relics. Detailed information for each of the relics was recorded, including dimensions (ranging from 12×22×7.5 cm to 30×45×2.5 cm) and preservation status, to ensure consistency in experimental design and result analysis. A subset of the dataset is shown in Fig. 1

Fig. 1. Relics from the Zhejiang museum

Category I (Simple Structure Type): Paper-based books with single-pattern flat carvings (*e.g.*, floral patterns). These relics feature clear outlines and uniformly distributed fibers.

Category II (Moderate Complexity Type): Basic root carvings with multi-layer reliefs (*e.g.*, landscape scenes with depth layers). These relics exhibit moderate texture variations and local occlusions.

Category III (High Complexity Type): Complex root carvings with sculpture in the round (*e.g.*, figure sculptures with curved surfaces). These relics have intricate fiber interweaving and significant depth variations.

Image Acquisition Solution

To prevent physical damage to root carvings and paper-based relics while reducing costs, this study employed a consumer-grade smartphone for data acquisition, thereby, eliminating the need for professional lighting or scanning equipment.

All images were captured at a uniform resolution of 1920×1080. Camera settings included a focal length range of 24 to 120 mm, ISO between 100 and 200, a shutter speed

of 1/100s, automatic white balance, and no flash. A handheld gimbal (DJI OM 6) was used throughout the capture process to minimize shake and ensure image clarity.

A two-dimensional acquisition scheme-combining 360° horizontal rotation with $\pm$ 30° pitch adjustments was employed for comprehensive coverage. The camera completed a full 360° horizontal rotation around each of the relics. At each horizontal position, the camera pitch was adjusted to three discrete angles: -30°, 0°, and +30°. The -30° pitch was used to capture details of the bottom and recessed areas, the 0° pitch for the main body, and the +30° pitch for the top and overhanging structures, ensuring all surfaces of the relics were thoroughly documented. The shooting setup is shown in Fig. 2.

The smartphone was orbited 360° around the relics to ensure full angular coverage and capture details that may be easily overlooked on the top and bottom regions of the relics. For each of the relics, a 40 to 60 second video was recorded at 30 frames per second (FPS); processing this video offers an efficient alternative to the labor-intensive process of capturing hundreds of individual static images. Following acquisition, FFmpeg was used to extract frames at a sampling rate of 2 frames per second, resulting in 80 to 120 images per relics.
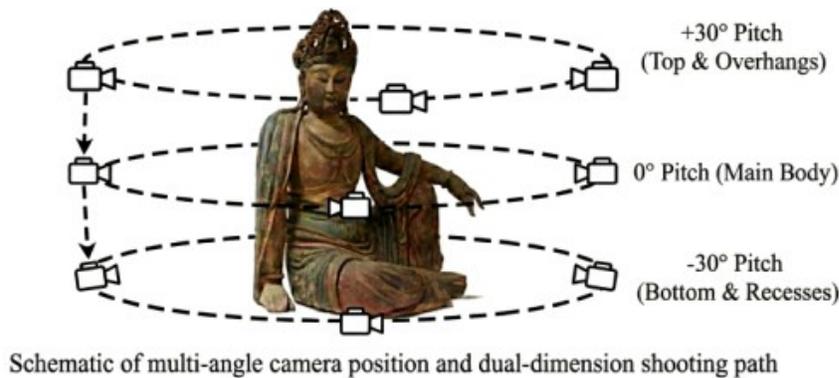

Schematic of multi-angle camera position and dual-dimension shooting path

Fig. 2. Schematic diagram of image acquisition scheme

Subsequently, OpenCV was used to preprocess the images, with the following operations: white balance was adjusted through color correction to preserve the authentic color fidelity of paper and root carvings, mitigating color casts such as yellowing; a Gaussian blur filter (with a kernel size of 3×3) was applied for noise reduction, which reduces sensor noise while minimizing the loss of fine fiber details; and the Region of Interest was centered on the main body of the relics to lower the computational complexity of subsequent steps.

Reconstruction Process Based on 3DGS

The reconstruction workflow is illustrated in Fig. 3. First, an image frame extraction operation was performed to extract images from video data at a specified frame rate. Multi-view images were used as input, with 45 to 60 images employed for reconstructing each of the relics.

The Structure from Motion (SfM) algorithm in COLMAP was applied to estimate camera parameters. The sparse point cloud generated by the algorithm and the estimated camera poses were combined as input to initialize the Gaussian distribution. Finally, a radiance field representation was optimized, and the 3D model was generated.

Figure 4 overviews the complete 3D reconstruction pipeline from image acquisition to model rendering, while Fig. 5 details the core 3D Gaussian Splatting (3DGS) training

process, specifically including the initialization from Structure-from-Motion (SfM) sparse point clouds to the technical specifics of adaptive Gaussian density control and differentiable rendering.
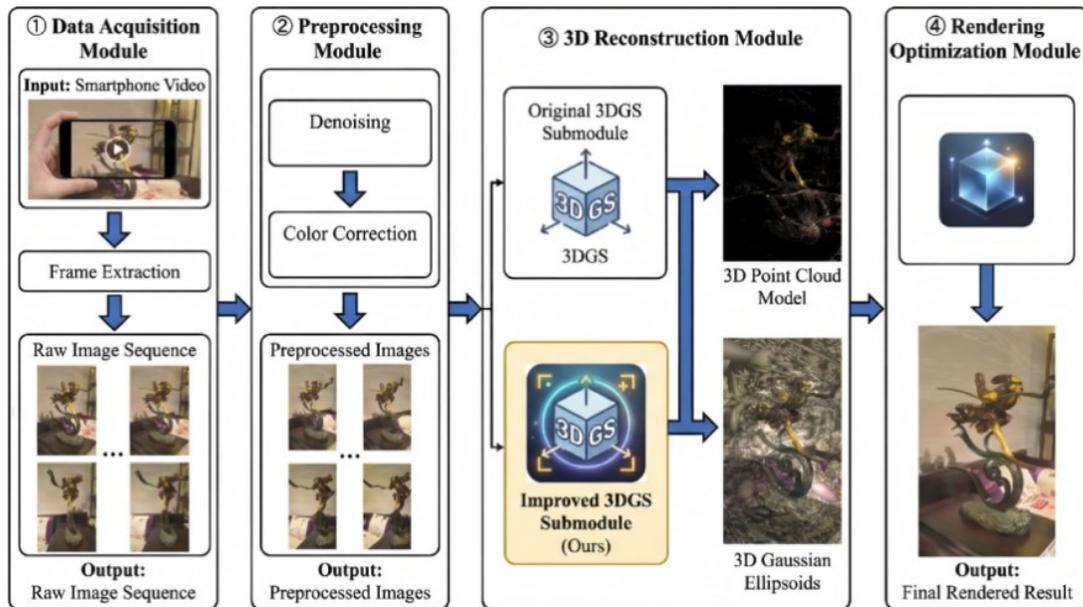

Fig. 3. 3D reconstruction process for cultural relic

Sparse reconstruction based on SfM

COLMAP was used for feature matching and camera pose estimation, with preprocessed images fed into the pipeline for processing. The Scale-Invariant Feature Transform (SIFT) algorithm was applied to detect keypoints, extracting at least 2,000 feature points from each image to improve the matching accuracy of fine paper textures.

The Fast Library for Approximate Nearest Neighbors (FLANN) matcher was utilized to establish feature correspondences between images. A ratio test with a threshold set to 0.8 was performed to filter out false matches, thereby ensuring the accuracy of feature pairs. Through Structure from Motion (SfM), the intrinsic parameters (focal length, principal point) and extrinsic parameters (position, pose) of each camera were estimated, generating a sparse point cloud with an average density of 5,000 to 8,000 points per relic. Finally, the Levenberg-Marquardt algorithm was employed for bundle adjustment to optimize camera poses and point coordinates.

The original 3D Gaussian Splatting (3DGS) method relies solely on a fixed threshold of pixel reconstruction error as its trigger condition, failing to account for scene material properties. The present method introduces a dual-trigger mechanism that integrates both pixel reconstruction error and bio-fiber texture gradient value. Densification is activated when the reconstruction error exceeds 0.05 and the texture gradient value is greater than 20. Conversely, pruning is triggered when the reconstruction error falls below 0.01 and the texture gradient value is less than 5 — conditions corresponding to texture-poor regions.

This approach is tailored to the textural characteristics of bio-materials. Thresholds are dynamically configured: a Gaussian cloning upper limit of 500,000 is set for densely fibrous regions, while an opacity pruning threshold of 0.1 is applied for non-textured areas. Furthermore, these thresholds are adaptively adjusted based on the relics's complexity level

(simple, medium, or high), thereby enhancing the reconstruction accuracy of fine fibrous details.

Building upon the texture gradient characteristics of bio-materials, the proposed method introduces a "region-adaptive" strategy for density control. Unlike the original 3DGS framework, which employs a "globally uniform" densification and pruning logic across all regions, the present approach enables density control to precisely align with the distribution patterns of bio-fibers. It prioritizes densification in regions with dense bio-fibers (*e.g.*, carved indentations in root carvings or interwoven fiber zones in paper) to preserve fine details, while applying "aggressive pruning" in texture-poor areas to eliminate redundant Gaussians. This strategy effectively balances reconstruction accuracy with computational efficiency.

3D Gaussian initialization

3DGS utilizes the sparse 3D point clouds and camera poses derived from Structure-from-Motion (SfM) as initial inputs, initializing each 3D point as a Gaussian distribution.

These Gaussian distributions are characterized by parameters including position (mean), covariance matrix, and transparency. A Gaussian function G(x) is used to represent a 3D point in space. G(x) is characterized by a central point position μ and a covariance matrix Σ, and its definition at a given position is as follows.

$$G(\mathbf{x}) = \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^{T}\boldsymbol{\Sigma}(\mathbf{x} - \boldsymbol{\mu}) \right]$$

(1)

The covariance matrix Σ defines the shape and orientation of the ellipsoid at a given position, which possesses physical meaning only when Σ is positive semi-definite. However, this condition cannot be consistently guaranteed during the optimization process. To address this issue, positive definiteness is maintained by decomposing Σ (Eq. 2) throughout optimization,

$$\boldsymbol{\Sigma} = \mathbf{R}\mathbf{S}\mathbf{S}^{T}\mathbf{R}^{T}$$

(2)

where R is the rotation matrix of the Gaussian ellipsoid, and S is the scaling matrix for its principal axes.

In the world coordinate system, a 3D Gaussian distribution is defined by a full 3D covariance matrix Σ. However, projecting the 3DGS distribution onto a 2D plane (*i.e.*, transforming from perspective projection to orthographic projection) constitutes a nonlinear transformation. To avoid introducing non-affine distortion during the projection transformation, a Jacobian matrix is introduced to perform local linear approximation of the nonlinear transformation, converting Σ into Σ' in-camera coordinates, as shown in Eq. 3,

$$\boldsymbol{\Sigma} = JW\boldsymbol{\Sigma}W^{T}J^{T}$$

(3)

where W is the transformation matrix from the world coordinate system to the camera coordinate system.

To obtain the rendered image, in addition to defining the fundamental attributes of the 3DGS (Eqs. 1 and 2), it is necessary to incorporate color attributes for realistic rendering. Each Gaussian is described by opacity α and color C defined by spherical harmonic (SH) coefficients. The color C of a specific pixel is calculated using N ordered Gaussian points-within a certain radius of the point cloud-that influence the pixel, with the calculation method detailed in Eq. 4,

$$C = \sum_{i \in N} c_i \alpha_i \prod_{j=1}^{i-1} (1 - \alpha_j)$$

$$(4)$$

where α_i is the opacity of the current point i, c_i is the color of the current point i, and α_j is the opacity of the preceding points.

After image synthesis, the loss is computed based on the difference between the rendered image and the input ground-truth image. The loss function L is composed of an L_1 loss and a structural similarity loss$L_{D\text{-SSIM}}$, namely,

$$L = (1 - \lambda)L_1 + \lambda L_{D\text{-SSIM}}$$

$$(5)$$

where λ is the weight parameter, the L_1 loss measures the absolute difference in pixel colors, and the $L_{D\text{-SSIM}}$ loss measures the similarity between pixel windows.

To represent the geometric structure of the scene, 3DGS adaptively modifies the distribution of the point cloud based on gradients to generate new Gaussians. Due to the low quality of the initial point cloud, continuous optimization is required. To mitigate relics potentially caused by points positioned too close to the camera, points with opacity below a set threshold or located excessively near the camera are pruned during each training iteration. When gradients exceed a predefined threshold, over-reconstruction or under-reconstruction may occur; cloning operations are applied to under-reconstructed regions, while split operations are performed on over-reconstructed regions. By periodically densifying and pruning Gaussians, the number of 3D Gaussians is controlled, allowing for adaptive convergence to an optimal density that balances computational efficiency and scene detail fidelity, thereby achieving superior 3D scene representation.

Adaptive optimization based on the characteristics of root carvings and paper-based cultural relics

Two enhancements are implemented based on the 3DGS optimization process to address the unique characteristics of root carvings and paper-based cultural relics.

First, a gradient-based adaptive density control is adopted: for regions with high gradient magnitudes (*e.g.*, carved edges, fiber intersections), additional Gaussian volumes are cloned (with a maximum total of 500,000) to capture fine details; for low-gradient regions (*e.g.*, flat paper surfaces), Gaussian volumes with opacity lower than 0.1 are pruned to reduce redundancy and computational load.

Second, a translucency-aware color loss function is introduced. Considering that the light-scattering property of paper fibers causes subtle color variations in overlapping regions, the translucency model is incorporated into the color loss function. This loss function combines L1 loss (to ensure color accuracy) and a translucency term (to preserve fiber textures), with weight parameters $\lambda_1 = 0.7$ and $\lambda_2 = 0.3$ (determined *via* cross-validation) to balance color fidelity and detail preservation.

The model training involves a total of 30,000 iterations, with an initial learning rate of 0.00016. Adjustments are made at regular intervals to ensure stable convergence, balancing the reconstruction speed and detail quality for paper-based cultural relics.

Differentiable rendering

An optimized 3DGS volume is rendered into a 2D image using a tile-based rasterizer. First, the optimized camera pose is used to project each Gaussian volume from 3D world coordinates to 2D screen space. Subsequently, the image is partitioned into 32×32 tiles, and the Gaussian volumes within each tile are sorted by depth to resolve

bioresources.cnr.ncsu.edu

occlusions. Finally, the pixel color is calculated by accumulating the contribution values of all Gaussian volumes covering the pixel. In this process, the product term accounts for the transparency of preceding Gaussian volumes, simulating the light scattering effect.

The semi-transparency term

Traditional alpha-blending algorithms, which rely on linear superposition for color fusion in rendering, struggle to accurately model the translucent properties of paper-based relics. The translucency of paper-based relics (such as ancient Xuan paper and historical book pages) is determined by their fibrous structure, resulting in nonlinear characteristics in light transmission and scattering. Additionally, these relics often exhibit layering and ink diffusion, leading to significant regional variations in translucency. To precisely characterize these properties, authentically render the transparent quality and fibrous layering of paper, and meet the core requirements of authenticity and precision in 3D reconstruction of paper-based relics, this study specifically designed and introduced an independent translucency term. This module is constructed based on the physical principles of light transmission in paper media, tailored to the practical needs of attribute representation in relics reconstruction.

Paper is a porous fibrous translucent medium with distinct physical characteristics in light transmission and scattering. The newly introduced independent translucency term can directly constrain the translucency-related parameters of 3D Gaussians. This ensures that the optimization of Gaussian properties strictly adheres to the inherent light propagation laws of paper, enabling accurate characterization of its translucency. If the constraint for paper translucency were only embedded within the color reconstruction loss function and indirectly optimized via blending algorithms, the model might sacrifice the physical plausibility of translucency to fit pixel colors. This would ultimately lead to distorted transparency and confused translucency features in overlapping areas within the reconstruction, severely undermining both the academic research value and practical application value of the results.

By introducing an independent translucency term, this study achieves a decoupled optimization of the physical constraints for paper translucency and pixel-level color fusion. The translucency term ensures that the Gaussians accurately represent the differences and gradient features of translucency across various paper types and relics regions, while the blending algorithm guarantees the visual naturalness of the final rendered image. Their synergy effectively resolves issues such as color distortion and loss of detail layers inherent in traditional methods, significantly enhancing both the precision and the authenticity of the 3D reconstruction for paper-based relics.

Experimental Environment and Evaluation Indicators

Experimental environment

All experiments were conducted on a workstation with the following configurations: the operating system is Windows 11 Professional, the Central Processing Unit (CPU) is an Intel Core i9-13900K with 24 cores and a base clock speed of 3.0 GHz, and the Graphics Processing Unit (GPU) is an NVIDIA GeForce RTX 5080 equipped with 24 GB of video memory.

The software environment included PyTorch 2.0.1 (with GPU acceleration enabled *via* CUDA 12.1), COLMAP 3.8, FFmpeg 6.0, and OpenCV 4.8.0.

Evaluation indicators

To evaluate the performance of SLAM, this study employs Absolute Trajectory Error (ATE) and Root Mean Square Error (RMSE) to assess the accuracy of camera pose estimation. This metric quantifies localization error by computing the root mean square of absolute errors between the estimated trajectory and the ground-truth trajectory. Rendering quality is evaluated using Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS). 1) PSNR measures the similarity between rendered images and ground-truth images, with higher values indicating superior rendering quality. 2) SSIM evaluates structural similarity in rendered images, where higher values denote closer visual fidelity to real scenes. 3) LPIPS assesses perceptual differences between rendered and ground-truth images through deep neural network models. A lower value indicates that the perceptual effect is closer to that of the ground truth.

PSNR measures the quality of reconstructed images by calculating the Mean Squared Error (MSE) between the original image and the reconstructed image, as follows,

$$\text{PSNR} = 10 \cdot \lg\left(\frac{M_1^2}{M_S}\right) \tag{6}$$

where M_1 denotes the maximum value of the image pixels, and M_s represents the Mean Squared Error (MSE).

The Structural Similarity Index (SSIM) measures the perceptual similarity between the original and reconstructed images by comprehensively evaluating three key characteristics: luminance, contrast, and structural consistency. It is defined as follows,

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{7}$$

where the subscripts x and y denote the original image and reconstructed image respectively; μ_x and μ_y represent the mean luminance of the images; σ_x and σ_y indicate the standard deviations of the images; σ_{xy} signifies the covariance between the images; and $C_1, C_2,$, is used to prevent the denominator from being zero.

LPIPS (Learned Perceptual Image Patch Similarity), also referred to as perceptual loss, is a deep learning-based image similarity metric that quantifies the perceptual discrepancy between images in deep feature spaces. Its computational formulation is defined as follows,

$$\text{LPIPS}(x, y) = \sum_{l=1}^{L} \frac{1}{H_l W_l} \sum_{h,w}^{H_l, W_l} \left\| \mathbf{w}_l \odot \left(x_{hw}^l - y_{hw}^l\right)\right\|_2^2 \tag{8}$$

where L is the number of feature layers; l is the feature extraction layer; H_l and W_l are the height and width (in pixels) of the feature map at the l-th layer, respectively; w_l is the channel weight vector of the l-th layer; x_{hw}^l and y_{hw}^l are the features of the original image and the reconstructed image at the l-th layer and pixel position (h,w), respectively; and $\odot$ denotes the element-wise multiplication.

Higher values of PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index Measure), coupled with lower values of LPIPS (Learned Perceptual Image Patch Similarity), indicate superior reconstruction quality. Conversely, deviations from this pattern suggest challenges to converge effectively during model optimization.

RESULTS AND DISCUSSION

The experimental results are presented in Tables 1 and 2. For Category I cultural relics (flat carvings), the proposed 3DGS framework demonstrated significantly better reconstruction performance than Nerf and Instant Neural Graphics Primitives (Instant-NGP). The 3D models accurately preserved the uniform fiber distribution of flat paper without texture blurring. This result was attributed to the framework's optimized Gaussian initialization strategy (using small covariance in detailed areas). For Category II cultural relics (multi-layer reliefs), the 3DGS framework maintained excellent performance. In comparison, NeRF exhibited blurred inter-layer boundaries due to its limited capability in modeling depth variations. Although Instant-NGP improved depth representation, it suffered from reduced color saturation in overlapping relief layers. For Category III cultural relics (such as high-complexity figure sculptures), the 3D models generated by the proposed framework accurately captured the curved contours of human costumes and the complex fiber interweaving patterns in folded areas. This is attributed to the framework's Gaussian splitting/cloning operations, which adaptively increase Gaussian density on high-gradient surfaces. In contrast, NeRF failed to distinguish subtle depth changes of cultural relics, while Instant-NGP experienced loss of fiber details in occluded areas. The study included a 3DGS reconstruction framework specifically designed for root carving and paper-based cultural relics, which outperformed both NeRF and Instant-NGP in all metrics. Compared with NeRF, it increased the average PSNR by 39.60%, SSIM by 65.84%, and decreased LPIPS by 90.57%. Relative to Instant-NGP, the framework still achieved an increase of 31.56% in PSNR, 42.58% in SSIM, and a decrease of 89.21% in LPIPS.

To validate the reconstruction effectiveness of the proposed method under various conditions, four distinct case studies are presented in Fig. 4:

The first case features a figurative root carving, where both the figurative form and the base texture are distinctly recognizable.

The second case utilizes a scroll painting with high transparency.

The third case involves a bio-material relic captured in an environment with blurry lighting and prominent specular reflections.

The fourth case selected a dark-colored root carving ornament characterized by low luminance and complex texturing.

The NeRF-based processing introduced noticeable artifacts: the background appeared blurred with smoke-like noise and color distortion, the wall surface became mottled, and object edges were blurred, resulting in an overall loss of material fidelity. In contrast, the proposed 3D Gaussian Splatting (3DGS) reconstruction successfully preserved the color and textural details of both background and foreground with high fidelity, delivering a clearer and more natural visual representation.

Among all categories of cultural relics, the proposed 3DGS framework reduced the average reconstruction time by 68.9% compared with NeRF and by 48.2% compared with Instant-NGP. This efficiency advantage is crucial to the workflow of museums, as it enables the large-scale and rapid digital processing of root carving and paper-based cultural relics.
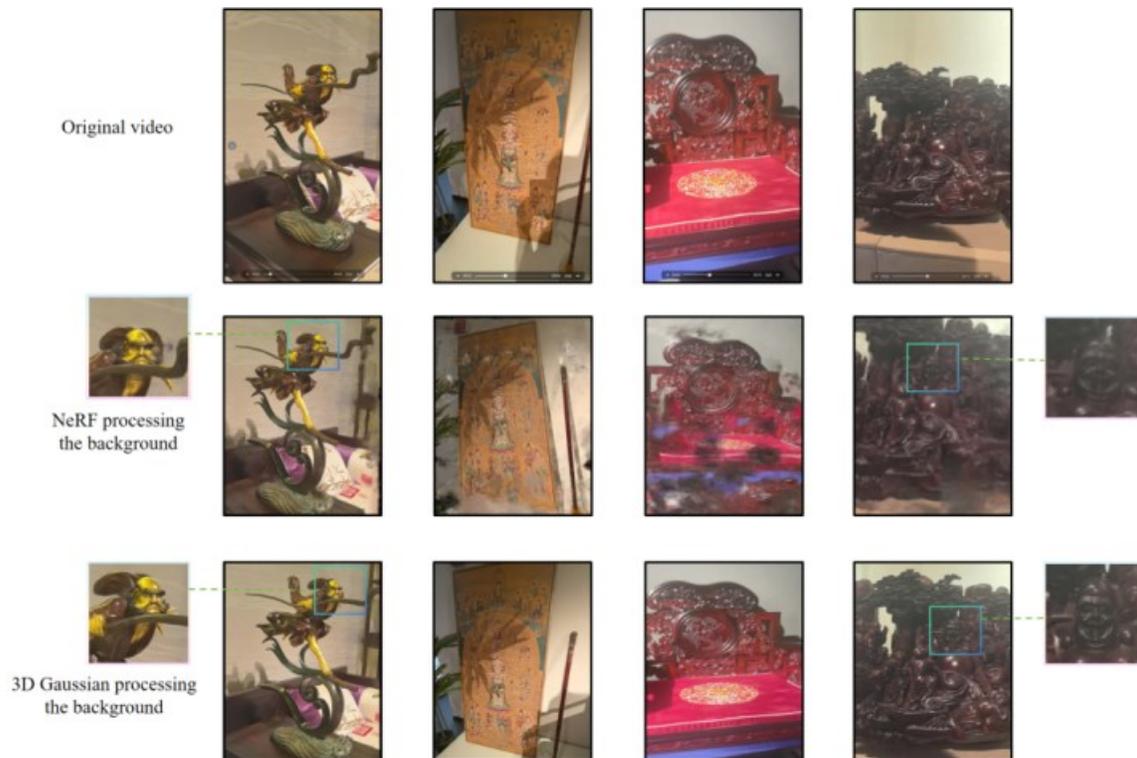
Fig. 4. Reconstruction effects of ancient cultural relics of varying complexity

The proposed 3DGS-driven framework addresses the core challenges in digitizing root carvings and paper-based relics. It uses a consumer-grade smartphone (costing approximately $1,000) to replace laser scanners (priced between $50,000 and $200,000), thereby reducing equipment costs by over 98%. Non-contact image acquisition eliminates the risk of physical damage to the fragile paper fibers.

Compared to NeRF, the framework significantly improves the preservation of the appearance of microfibers through adaptive Gaussian optimization and a translucency-aware loss function, thereby resolving the blurriness issue of NeRF in capturing the fine details of relics. The core of this improvement lies in 3DGS's representation of scenes as discrete Gaussian volumes rather than continuous radiance fields, which enables accurate modeling of the microfiber structures of paper.

The average reconstruction time per relic was 14 minutes, meeting the workflow requirements for large-scale digitization in museums. The efficiency improvement stems from the tile-based rasterization technology of 3DGS, which supports GPU parallel acceleration.

bioresources.cnr.ncsu.edu

Table 1. Comparison of Reconstruction Quality and Reconstruction Time for Cultural Relics of Different Complexity Types

Level of Complexity	Model	Peak signal-to-noise ratio (PSNR)/db	Structural similarity index measure(SSIM)	Learned perceptual image patch similarity(LPIPS)
Category I (Simple Structure Type)	Nerf	28.16	0.640	0.148
	Instant-ngp	29.14	0.682	0.122
	3DGS	34.14	0.884	0.012
Category II (Moderate Complexity Type)	Nerf	24.31	0.523	0.156
	Instant-ngp	25.64	0.654	0.133
	3DGS	33.84	0.872	0.016
Category III (High Complexity Type)	Nerf	20.19	0.417	0.174
	Instant-ngp	22.32	0.503	0.162
	3DGS	33.45	0.865	0.018
Mean	Nerf	24.22	0.527	0.159
	Instant-ngp	25.7	0.613	0.139
	3DGS	33.81	0.874	0.015

Table 2. Comparison of Reconstruction Time Across Different Models

Algorithm model	Average time consumed/min
3DGS	14
Nerf	45
Instant-ngp	27

To systematically evaluate the contribution of key modules in the proposed framework, the following four experimental groups were designed for comparison:

The Full Model incorporates the complete 3DGS framework with both adaptive density control and the translucency-aware color loss function.

Ablation Group A removes the adaptive density control module.

Ablation Group B removes the translucency term from the color loss function.

The Baseline Model uses the standard 3DGS pipeline without any of our proposed optimization strategies.

The experiments were conducted on the ancient Xuan paper dataset, with specific results shown in Table 3.

After removing the adaptive density control module (Ablation Group A), all rendering quality metrics declined significantly, showing a substantial perceptual quality gap compared to the Full Model. The adaptive density control module played a decisive role in accurately reconstructing the complex geometric forms on the relic surfaces. Visually, the reconstruction results of Ablation Group A appeared blurred in detailed regions and failed to clearly render edges. The module was key to achieving high-fidelity geometric reconstruction by adaptively cloning and splitting Gaussian ellipsoids in high-gradient regions (*e.g.*, edges and texture boundaries).

Removing the translucency term (Ablation Group B) led to slight decreases in PSNR and SSIM, but a more pronounced deterioration in LPIPS (+56.3%). This indicates that, although the absolute color error remained relatively low, the model's ability to simulate visual appearance characteristics—such as paper translucency and light-scattering effects—was impaired, reducing the perceived "realism" of the reconstruction. Unnatural color patches tend to appear in overlapping paper areas, failing to reproduce the soft color transitions and sense of depth that result from light penetrating multiple fiber layers in real

bioresources.cnr.ncsu.edu

paper. The translucency-aware loss function significantly enhanced the visual realism of rendered results by simulating the physical interaction between light and paper fibers.

The Baseline Model performed worst across all metrics, fully demonstrating the necessity of the two proposed modules. The Full Model achieved the best overall performance, showing that adaptive density control and the translucency-aware loss are complementary and synergistic modules that collectively enable high-quality reconstruction—from geometric details to visual appearance. This synergistic optimization is ultimately reflected in the excellent reconstruction results for biomass-type cultural relics.

Table 3. Quantitative Results of the Ablation Study

Experimental Group	PSNR (dB)	SSIM	LPIPS
Full Model	**33.84**	**0.872**	**0.016**
Ablation Group A (-ADC)	30.15 (-10.9%)	0.801 (-8.1%)	0.039 (+143.8%)
Ablation Group B (-ST Loss)	32.57 (-3.8%)	0.841 (-3.6%)	0.025 (+56.3%)
Baseline Model (Baseline 3DGS)	28.92 (-14.5%)	0.763 (-12.5%)	0.048 (+200.0%)

CONCLUSIONS

This study successfully validated the feasibility and application potential of a low-cost 3D Gaussian reconstruction workflow based on consumer-grade smartphones. The proposed framework can be widely applied to the digital preservation of fragile paper-based relics, offering not only a technical pathway characterized by low cost, high efficiency, and non-contact operation but also fostering a practical bridge between conservation science, humanities research, and public education. It thereby provides a demonstrative example for the sustainable dissemination and innovative utilization of cultural heritage.

1. The study introduced a 3DGS-driven framework specifically designed for root carving and paper-based cultural relics. Based on an acquisition scheme using consumer-grade smartphones, it achieved full-coverage image capture of fragile root carving and paper-based cultural relics. Compared with laser scanning, it reduced equipment costs by more than 98% and caused no physical damage. Compared with NeRF, it increased the average PSNR by 39.60%, SSIM by 65.84%, and decreased LPIPS by 90.57%. Compared with Instant-NGP, the framework still achieved an increase of 31.56% in PSNR, 42.58% in SSIM, and a decrease of 89.21% in LPIPS.

2. By enhancing 3DGS optimization (including adaptive density control and translucency-aware loss), the framework achieved high detail fidelity, meeting the scientific requirements for digital archiving and restoration of root carvings and paper-based relics. Based on the strategy of 3DGS for detail preservation-realized through adaptive density control and pruning of Gaussian volumes with low opacity-the covariance parameters were adjusted for the microfiber structure of paper (set to 0.5 mm in detail regions), addressing the issues of fiber blurriness in NeRF and partial fiber loss in Instant-NGP.

3. The explicit Gaussian rendering of 3DGS can more effectively suppress background noise (such as transient objects and uneven lighting), while the MLP ray sampling of NeRF is susceptible to interference; this robustness is crucial for museum applications. This framework provides a cost-effective and efficient technical solution for the digital conservation of paper-based cultural heritage, offering key support for the digital preservation of fragile paper-based relics.

ACKNOWLEDGMENTS

REFERENCES CITED

Barron, J. T., Mildenhall, B., Tancik, M., Hedman, P., Martin-Brualla, R., and Srinivasan, P. P. (2021). "Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields," in: *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, QC, Canada, 2021, pp. 5835-5844. https://doi.org/10.1109/ICCV48922.2021.00580

Barron, J. T., Mildenhall, B., Verbin, D., Srinivasan, P. P., and Hedman, P. (2021). "Mip-nerf 360: Unbounded anti-aliased neural radiance field," in: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, USA, 2022, pp. 5460-5469. https://doi.org/10.1109/CVPR52688.2022.00539

Cai, Z., Yang, J., Wang, T., Huang, H., and Guo, Y. (2024). "3D Reconstruction of buildings based on 3D Gaussian splatting," *The International Archives of the Photogrammetry*, *Remote Sensing and Spatial Information Sciences* 48, 37-43. https://doi.org/10.5194/isprs-archives-XLVIII-4-W10-2024-37-2024

Chen, Y., and Lee, G. H. (2024). "Dogs: Distributed-oriented Gaussian splatting for large-scale 3D reconstruction *via* Gaussian consensus," *Advances in Neural Information Processing Systems* 37, 34487-34512.

Dell'Unto, N., Forte, M., and Remondino, F. (2018). "3D documentation of cultural heritage: A review of best practices and technological insights," *Journal of Cultural Heritage* 31, 119-131.

Garbin, S. J., Kowalski, M., Johnson, M., Shotton, J., and Valentin, J. (2021). "Fastnerf: High-fidelity neural rendering at 200fps," *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, QC, Canada, 2021, pp. 14326-14335. https://doi.org/10.1109/ICCV48922.2021.01408

Gomes, L., Bellon, O. R. P., and Silva, L. (2014). "3D reconstruction methods for digital preservation of cultural heritage: A survey," *Pattern Recognition Letter* 50, 3-14. https://doi.org/10.1016/j.patrec.2014.03.023

Jebur, A. K. (2022). "The techniques of cultural heritage: Literature review," *Saudi J. Civ. Eng* 6, 108-114. https://doi.org/10.36348/sjce.2022.v06i04.006

Moyano, J., Cabrera-Revuelta, E., Nieto-Julián, J. E., Fernández-Alconchel, M., and Fernández-Valderrama, P. (2023). "Evaluation of geometric data registration of small objects from non-invasive techniques: Applicability to the HBIM field," *Sensor*s

23(3), article 33. https://doi.org/10.3390/s23031730

Müller, T., Evans, A., Schied, C., and Keller, A. (2022). "Instant neural graphics primitives with a multiresolution hash encoding," *ACM Transactions on Graphics (TOG)* 41(4), 1-15. https://doi.org/10.48550/arXiv.2201.05989

Munoz-Pandiella, I., Bosch, C., Guardia, M., Cayuela, B., Pogliani, P., Bordi, G., and Charalambous, P. (2024). "Digital 3D models for medieval heritage: Diachronic analysis and documentation of its architecture and paintings," *Personal and Ubiquitous Computing* 28(3), 521-547. https://doi.org/10.1007/s00779-024-01816-6

Nobuhara, S., and Matsuyama, T. (2003). "Dynamic 3D shape from multi-viewpoint images using deformable mesh model," in: *3rd International Symposium on Image and Signal Processing and Analysis*, Rome, Italy, 2003, pp. 192-197. https://doi.org/10.1109/ISPA.2003.1296892

Rhodin, H., Robertini, N., Richardt, C., Seidel, H. P., and Theobalt, C. (2015). "A versatile scene model with differentiable visibility applied to generative pose estimation," in: *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, pp. 765-773. https://doi.org/10.1109/ICCV.2015.94

Ryan, C., Haist, T., Laskin, G., Schrder, S., and Reichelt, S. (2024). "Technology selection for inline topography measurement with rover-borne laser spectrometers," *Sensors* 24(9), article 30. https://doi.org/10.3390/s24092872

Samavati, T., and Soryani, M. (2023). "Deep learning-based 3D reconstruction: A survey," *Artificial Intelligence Review* 56(9), 9175-9219. https://doi.org/10.1007/s10462-023-10399-2

Turki, H., Ramanan, D., and Satyanarayanan, M. (2021). "Mega-nerf: Scalable construction of large-scale nerfs for virtual fly-throughs," in: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, USA, 2022, pp. 12912-12921. https://doi.org/10.1109/CVPR52688.2022.01258

Wang, H., Ren, J., Huang, Z., Olszewski, K., Chai, M., Fu, Y., and Tulyakov, S. (2022). "R2l: Distilling neural radiance field to neural light field for efficient novel view synthesis," in*: European Conference on Computer Vision*, Tel Aviv, Israel, pp. 612-629. https://doi.org/10.48550/arXiv.2203.17261

Wang, W. (2024). "Real-time fast 3D reconstruction of heritage buildings based on 3D Gaussian splashing," in: *2024 IEEE 2nd International Conference on Sensors, Electronics and Computer Engineering (ICSECE)*, Jinzhou, China, pp. 1014-1018. https://doi.org/10.1109/ICSECE61636.2024.10729491

Yang, Z., Gao, X., Zhou, W., Jiao, S., and Zhang, Y. (2024). "Deformable 3D Gaussians for high-fidelity monocular dynamic scene reconstruction," in: *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 2024, pp. 20331-20341. https://doi.org/10.1109/CVPR52733.2024.01922

Yu, Z., Chen, A., Huang, B., Sattler, T., and Geiger, A. (2024). "Mip-splatting: Alias-free 3D Gaussian splatting," in: *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 2024, pp. 19447-19456. https://doi.org/10.1109/CVPR52733.2024.01839

Zhang, K., Riegler, G., Snavely, N., and Koltun, V. (2020). "NeRF++: Analyzing and improving neural radiance fields," *ArXiv*. https://doi.org/10.48550/arXiv.2010.07492

Zhao, S., Kang, F., Li, J., and Ma, C. (2021). "Structural health monitoring and inspection of dams based on UAV photogrammetry with image 3D reconstruction," *Automation in Construction* 130. https://doi.org/10.1016/j.autcon.2021.103832

Zheng, P., Wei, L., Jiang, D., and Zhang, J. (2025). "3D Gaussian splatting against moving objects for high-fidelity street scene reconstruction," *ArXiv* abs/2503.12001. https://doi.org/10.48550/arXiv.2503.12001